FLOOD PREDICTION USING SUPPORT VECTOR MACHINE ALGORITHM ON MOBILE APPLICATION

***Afif Shyamsul Syahiran Suhaimi*[1] *, Habibah Ismail*[2*]*, and Ismail Ahmedy*[3]**

[1]College of Computing, Informatics and Mathematics
Universiti Teknologi MARA
Terengganu, Malaysia

[2]Centre of Foundation Studies
Universiti Teknologi MARA
Selangor, Malaysia

[3]Faculty of Computer Science and Information Technology
Universiti Malaya
Kuala Lumpur, Malaysia

Emails: afifshyamsul12@gmail.com[1], habibahismail@uitm.edu.my[2*], ismailahmedy@um.edu.my[3]

ABSTRACT

Floods are recurrent natural disasters that can have a devastating impact on societies, economies, and the environment. Therefore, it is critical to predict and manage flood situations promptly to minimize the damage they cause. However, many people are unaware of flood risks, and there are limited mobile applications that can provide timely and accurate flood predictions. This study explores the application of the Support Vector Machine (SVM) algorithm for flood prediction in a mobile application. The aim is to provide users with timely and accurate flood predictions, enabling them to make informed decisions and take necessary precautions to mitigate flood impacts. By integrating the SVM algorithm into a mobile application, users gain convenient access to flood predictions, empowering them to be better prepared for potential flooding events. The user-friendly platform delivers critical flood forecasts, ensuring individuals and communities can respond effectively to flood situations. The evaluation of the SVM algorithm's performance reveals an achieved accuracy of 66.66%. In conclusion, this study underscores the potential of the SVM algorithm for flood prediction in a mobile application. These findings contribute to the field of flood forecasting technology, paving the way for more sophisticated and effective flood prediction tools in the future.

Keywords: Floods; Support Vector Machine; Mobile application; Accuracy; Mitigation.

1.0 INTRODUCTION

Floods are a natural disaster that regularly happens in Malaysia, as they happen annually [1]. A flood occurs when water that has been accumulating in a river suddenly flows out of the river and overtops the usually dry land. There are many types of floods; flash floods, tidal floods, and monsoon floods. Flood disasters can be severe enough to wipe out entire regions due to their size and force. They cause significant losses, including loss of life, damage to property and food supplies, and destruction of government infrastructure.

The floods of December 2021 in Malaysia left almost 50 dead, required the evacuation of about 400,000 people, and resulted in an overall estimate of RM6.1 billion in financial losses. Unprecedented volumes of rainfall left areas on the west coast of Peninsular Malaysia under almost four meters of water and turned roads into rivers [2]. Most victims are unaware of the upcoming flood disaster, as there was no specific flood prediction system. In other instances, although authorities had already warned residents about the flood disaster, many disregarded the warnings due to a lack of detailed information about flood scenarios, the timing of the floods, and the extent of inundation.

Existing monitoring and warning technologies for flooding can be categorized into two groups; remote sensing data and local sensing data. Remote sensing data typically utilize satellites to capture cloud images and predict rain occurrences. Local sensing data involve deploying sensor nodes along rivers to collect flood parameters within specific study regions, such as rivers. The flood monitoring and alert system study was conducted by a former University Technology of PETRONAS student, focusing solely on water level readings without considering other critical flood parameters. Consequently, the system design may not be comprehensive enough to effectively monitor and alert potential floods.

The flood prediction mobile application can assist users in predicting floods anywhere and anytime. This system can benefit both individuals who have experienced flooding and those who have not. It is accessible from any location and can be checked at any time, making it an invaluable tool for staying informed about flood risks. Moreover, it contributes significantly to flood-related studies and aids Malaysia's Civil Defence Force in their operations. The MCDF can utilize the system to pinpoint flood-prone areas and safely evacuate residents. Additionally, the system provides early flood warnings, enabling them to take preventive measures and minimize damage, which is crucial for reducing the overall impact of flooding. This system is capable of measuring various flood parameters, including humidity, temperature, and rainfall.

2.0 RELATED WORKS ON FLOOD PREDICTION

In recent years, research has focused on flood prediction to develop accurate and effective methods for predicting and mitigating flood damage. Support Vector Machines (SVM), Artificial Neural Networks (ANN), Long Short-term Memory (LSTM), Quadratic Support Vector Machine (Q-SVM), K-Nearest Neighbours (K-NN), and Linear Discriminant Analysis (LDA) have all been employed in these studies.

One such study, conducted by [3], analyzed data from 2005 and 2006 encompassing seven significant river floods in central Chiang Mai, Thailand. The study aimed to evaluate SVM's effectiveness in flood prediction. Results indicated that SVM models outperformed Multilayer Perceptron (MLP) models. Furthermore, results from blind test sets demonstrated that SVM models could effectively alert people to impending floods before they occur.

In 2022, [4] proposed a universal model for predicting streamflow in Peninsular Malaysia using Machine Learning (ML) models based on SVM, ANN, and Long Short-Term Memory (LSTM). The study found that the ANN model (specifically ANN3) could predict streamflow in Peninsular Malaysia's rivers the best. In another study referenced as [5], employing Linear Support Vector Machine, Quadratic Support Vector Machine, K-nearest neighbour, and Linear Discriminant Analysis, SVM was identified as the most accurate and precise method for numerical prediction.

Using the Internet of Things (IoT) and Feed-Forward Artificial Neural Network (FFANN), a study by [6] predicted heavy rains ahead of time to prevent deaths and property damage caused by flash floods. The flood forecasting system achieved 96% accuracy, 97% precision, a 95.9% recall rate, and an F1-Measure of 96.4%. Additionally, [7] utilized dynamic artificial neural networks (ANNs) to assess actual and fractional-storage. It was noted that ANNs exhibit variability in prediction accuracy when the same network topology is maintained. Moreover, [8] investigated flood possibilities in a region using IoT and artificial neural networks. With an 86% accuracy rate, the gradient descent method with adaptive learning demonstrated effective rain and flood prediction capabilities. Furthermore, [9] employed binary classification methods including Binary Logistic Regression, Support Vector Classifier (SVC), and K-Nearest Neighbours (KNN) to enhance flood prediction accuracy.

In conclusion, these studies have used various methods and algorithms to devise accurate and useful flood prediction models. They have been conducted in diverse locations using different types of data and have concluded that SVM, ANN, and LSTM are effective methods for predicting floods. Based on previous research, SVM emerges as the most suitable algorithm for further development due to its high accuracy and ability to handle diverse datasets, which results in superior prediction accuracies compared to other algorithms.

2.1 Floods

Flooding takes place when the land and vegetation are unable to absorb excess rainwater. It can also happen when a river's banks break, causing water to spill over onto the surrounding floodplain. Floods can result from natural events like hurricanes, weather patterns, and melting snow. Other types of floods, such as those caused by tsunamis and coastal surges, stem from natural occurrences like underwater earthquakes and the gravitational pull of the moon that affects high tides. In addition, various human activities can contribute to flooding [10].

Regional flooding happens when prolonged heavy rainfall falls over a vast region for days or even weeks, leading to rapid river rises that flood large areas and result in significant economic damage. Flash floods are also referred to as upstream floods, regional floods, and downstream floods [10]. The three main types of floods that impact numerous countries are flash floods, river floods, and coastal floods. Each of these flood types is explained in the following sections.

2.1.1 Flash Flood

Flash flooding is a type of flooding that happens quickly after a precipitation event, typically within six hours. It is often caused by intense rainfall and usually occurs near rivers or lakes, though it can also happen in areas without nearby water bodies [11].

Flash floods happen when the ground is overwhelmed and unable to absorb all the water from rainfall, and can also result from mudslides or the failure of dams or levees. Cities are especially susceptible to flash flooding due to the abundance of paved surfaces [11]. Fig. 1 depicts a flash flood in Kuantan, Pahang.

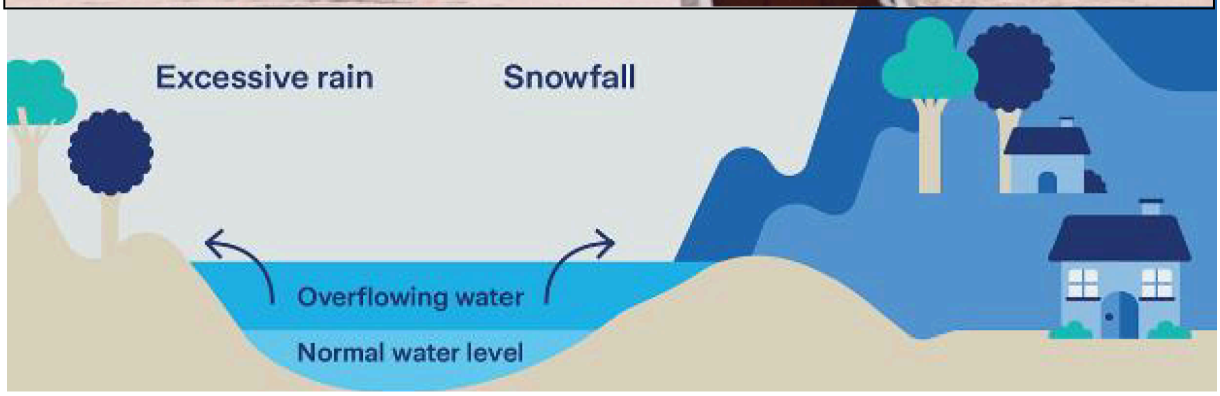


Fig. 2: How river flood occurred

The severity of a river flood is influenced by the topography, as well as the duration and intensity (volume) of rainfall in the river's catchment area. Additional factors include soil water saturation and the impact of climate change on rainfall duration and intensity. In flat areas, floodwaters rise more gradually and remain shallow, but they can persist for days. In contrast, floods in steep or mountainous regions can develop rapidly after heavy rain, drain quickly, and cause damage due to debris movement.

2.1.3 Coastal Flood

Coastal flooding occurs when seawater inundates previously dry land, a process driven by rising sea levels. This can happen in several ways, such as when water overflows a protective barrier, breaches a barrier, or floods directly. One form of coastal flooding is direct flooding, which happens when the land lies below sea level and lacks natural defenses like dunes. In such cases, seawater quickly spreads across the land [12].

Water can also spill over barriers, such as dunes or dams, when their height is exceeded. In this case, the excess water flows over and floods the land on the other side. Flooding can occur when waves are powerful enough to breach a natural or man-made barrier. If the barrier is broken or overwhelmed by these waves, it leads to coastal flooding. Understanding the causes and risks of coastal flooding is essential for preparedness and mitigation [12].

Several factors contribute to flooding along coastlines or nearby areas. The main contributors include the land's elevation relative to the sea, the extent of erosion and land subsidence, the loss of vegetation, and the occurrence of storm surges. Key causes of coastal flooding include:

- Low land elevation – Areas closer to sea level are more vulnerable to flooding, particularly from storm surges and high tides.
- Erosion and subsidence – The wearing away of land or its sinking can heighten the chances of flooding.
- Loss of vegetation – Natural protective barriers, like wetlands and mangroves, help shield coastlines. When these are destroyed, the land becomes more prone to flooding.
- Storm surges – Strong winds and low pressure from storms can drive water levels upward, pushing water inland and causing flooding.

Coastal flooding results in several challenges, including recurring road closures, diminished stormwater drainage efficiency, damage to infrastructure, and saltwater contamination of drinking water. These impacts can also have significant effects on public health. For example, when water systems are damaged and saltwater penetrates, the risk of exposure to harmful pathogens and chemicals rises [13].

2.2 Flood Impact

Floods can greatly affect people, ecosystems, and Malaysia as a whole. The extent of their impact often depends on variables such as the flood's location, duration, and intensity [14].

2.2.1 On Human

Floods can have both immediate and long-lasting effects on human health, impacting individuals directly and indirectly. The health of communities affected by flooding is particularly at risk, with disaster responders, healthcare workers, and essential service providers also being vulnerable [15]. In addition to drowning, which is often the most immediate health threat, floods pose a range of other dangers, including injuries, hypothermia, animal bites, infectious diseases, malnutrition, and mental health challenges.

While the economic losses and property damage from floods receive significant attention, the health impacts often intersect with these concerns in complex ways. Mental health issues, which can emerge in the aftermath of flooding, are frequently overlooked in comparison to acute physical health effects. However, it is now recognized that the psychological impact of flooding can persist for years, with anxiety being one of the most common disorders that arise. Other mental health issues often include depression, Post-Traumatic Stress Disorder (PTSD), insomnia, and psychosis.

2.2.2 The Environment

Flooding can cause significant environmental damage, including soil and bank erosion, bed erosion, silt accumulation, and landslides. It can also harm plant life, while contaminants carried by floodwaters may degrade water quality, disrupt ecosystems, and harm both plant and animal species. On the other hand, flooding can be beneficial to certain natural areas. Many wetland ecosystems depend on occasional flooding for their survival and can play a role in storing floodwaters, helping to reduce the risk of flooding in other regions [16].

In urban and rural areas, flooding can lead to the destruction of property. This can occur in both highly developed cities and more simple, rural settings. Essential services such as clean water, electricity, and transportation may be disrupted. In some cases, floodwaters can inundate homes, and in the worst instances, even wash them away [17].

Furthermore, floods negatively impact agriculture by causing oversaturation, soil fertility loss, and erosion, which damage crop fields, particularly those planted with winter crops. Additionally, flooding can contaminate groundwater, making it unsafe for drinking. The influx of harmful bacteria in floodwaters often leads to waterborne diseases, which further exacerbate health issues. Flooding also disrupts the natural water cycle, affecting freshwater ecosystems and aquatic life [18].

2.2.3 Malaysia

The floods that hit the country in late 2021 and early 2022 caused extensive damage to homes, vehicles, commercial properties, and both the manufacturing and agricultural sectors, as well as to public infrastructure. Total flood-related losses amounted to RM6.1 billion, roughly 0.40 percent of the country's nominal gross domestic product. As shown in Fig. 3, losses included RM1.6 billion in residential damage, RM0.5 billion in commercial property damage, RM1.0 billion in vehicle losses, RM90.6 million in agricultural damage, RM0.9 billion in manufacturing losses, and RM2.0 billion in damage to public infrastructure [2].

Fig. 3: Residential losses

The floods impacted 11 states across 60 districts, including Johor, Kelantan, Melaka, Negeri Sembilan, Pahang, Perak, Selangor, Terengganu, Sabah, Sarawak, and the Federal Territory of Kuala Lumpur.

2.3 Monitoring and Alert System

The process of monitoring refers to observing changes in a system and the flow of data, with the aim of detecting errors and supporting their resolution. It encompasses real-time processing, statistical analysis, and data examination, all of which are integral to information system monitoring. A monitoring system is made up of various software components that handle data collection, analysis, and presentation [19].

Alerting refers to the ability of a monitoring system to detect significant events that signify a major change in state and notify operators accordingly. This notification, called an alert, is typically a short message delivered via email, SMS, Instant Messaging (IM), or phone call. The alert is directed to the appropriate recipient, who is responsible for managing the event. Usually, the alert is logged as a ticket in an Issue Tracking System (ITS), also known as a ticketing system [19].

The condition of a system is assessed through the monitoring process, which can be approached proactively or reactively. Proactive monitoring involves tracking visible indicators like time series and dashboards and is often referred to as monitoring by administrators. Reactive monitoring, on the other hand, uses automated techniques to notify operators of important system changes, a process commonly known as alerting [19].

A flood monitoring and alert system is designed to track flood risks and issue alerts or notifications when specific conditions or thresholds are reached. These systems typically integrate sensors, communication networks, data management tools, and software to collect and analyze data, triggering alerts when flood risks are identified. Depending on the system, notifications can be sent via email, text messages, or mobile apps. These systems are commonly used by individuals, government bodies, and organizations to forecast floods and safeguard lives, property, and infrastructure.

2.4 Prediction/Forecasting

A prediction is essentially a forecast, but its scope extends beyond just weather. The prefix "pre" means "before," and "diction" relates to speech, so a prediction is a statement about the future. While it may be based on evidence or facts, it is often an educated guess [20]. Predictions are widely used in various fields, such as weather forecasting, finance, economics, and business demand forecasting. Additionally, predictions are common in areas like sports, politics, and many other domains.

Flood prediction is the process of estimating the probability and potential impact of future floods by utilizing historical data, statistical models, and various other tools. Such predictions are crucial for guiding evacuation efforts, emergency responses, and land-use planning, as well as for making informed decisions about infrastructure projects. To accurately predict floods, numerous types of data are required, including real-time rainfall, river level readings, storm characteristics, and detailed information about the river's drainage basin, such as soil moisture, ground temperature, snowpack, topography, vegetation, and the extent of impermeable surfaces. These variables are essential for estimating the size and severity of potential floods and improving forecast precision [21].

River flow forecasting is crucial for managing extreme hydrological events. As forecasting involves estimation, it is inherently challenging due to modeling errors, uncertainties, and the nonlinear nature of the processes involved. The objective of creating a flood early warning system is to reduce the risks and dangers posed by flooding and safeguard against its harmful effects [22]. Fig. 4 illustrates example of flowchart diagram used for prediction.

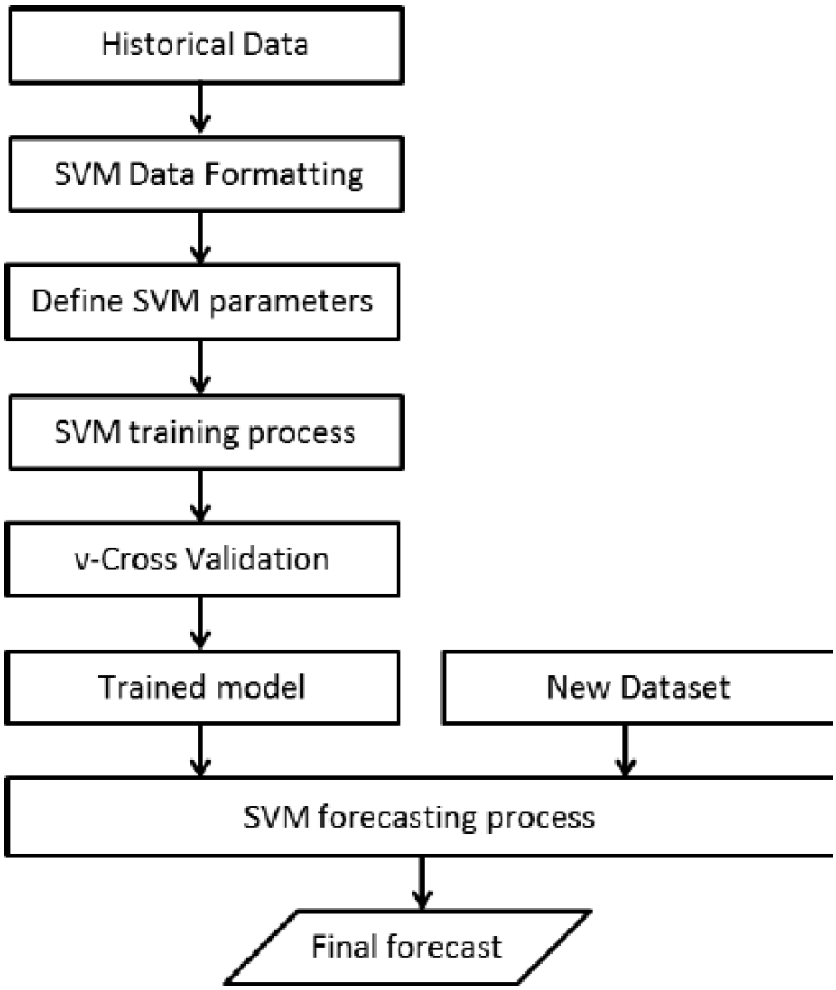


Fig. 4: Example of flowchart prediction in SVM

2.4.1 Application of Prediction/Forecasting for Flood in Real-Life

Flood prediction is vital for mitigating flood risks in Malaysia. A prime example is the impact-based flood forecasting and warning system developed for the Kelantan River Basin. This system combines information about potential flood impacts to improve both the precision and reliability of flood forecasts and warnings. By providing early warnings and insight into possible outcomes, the system helps authorities and communities prepare for and manage flood events more effectively.

A study by [1] shows that the impact-based flood forecasting and warning system has been successful in minimizing the effects of floods in the Kelantan River Basin. The research assesses the system's performance during the 2014 flood event and demonstrates that the system provided timely and accurate warnings, enabling communities and authorities to take appropriate preventive measures. The study also underscores the need for ongoing monitoring and enhancement of the system to maintain its effectiveness amid shifting weather patterns and other variables.

Overall, the impact-based flood forecasting and warning system in the Kelantan River Basin demonstrates the effective use of cutting-edge technologies and techniques in flood prediction. By factoring in potential impacts, the system enhances the ability of authorities and communities to prepare for and respond to floods, ultimately reducing the risks of damage and loss of life.

2.5 Machine Learning

According to IBM, machine learning is an area within artificial intelligence (AI) and computer science that focuses on using data and algorithms to replicate human learning, with continuous improvements in its accuracy.

Support Vector Machines (SVMs) are a type of machine learning algorithm used for classification and regression tasks. SVMs operate by identifying the hyperplane that best divides the data into two distinct classes. In the context of flood prediction, these classes would be "flood" and "no flood." SVMs have proven effective in predicting floods, as they can recognize patterns in historical data linked to flooding events. For instance, an SVM could be trained on past flood data from Malaysia, learning to identify correlations between variables like temperature, humidity, and rainfall that help predict the likelihood of a flood. Although the use of SVMs in flood prediction is still developing, it shows great promise as a tool for enhancing our understanding of floods and improving response strategies.

2.5.1 Support Vector Machine (SVM) Algorithm

SVM is a traditional machine learning technique primarily used for classification tasks. It operates by solving the dual form of high-dimensional problems, allowing the classifier to achieve structural risk minimization with the use of only a small number of support vectors [23]. SVM is a supervised machine learning algorithm designed for both classification and regression tasks. Its goal is to classify data points by finding a hyperplane in an N-dimensional space. The complexity of the hyperplane depends on the number of features; with two input features, the hyperplane is simply a line, with three features, it becomes a two-dimensional plane, and for more than three features, visualization becomes increasingly difficult [24].

The goal of the SVM algorithm is to find the optimal line or decision boundary that separates n-dimensional space into distinct classes, allowing for easy classification of future data points. This ideal boundary is known as a hyperplane [25].

The following are the advantages of employing SVM for classification and prediction:

- SVMs are effective in high-dimensional spaces, or when there are more features than training examples.
- SVMs are memory efficient, as they only use a subset of the training examples (support vectors) in the decision function.
- SVMs are versatile, as they can be used for both linear and non-linear classification, and they can also be used in regression tasks.
- The margin (the distance between the decision boundary and the closest training examples) of an SVM can be maximized, which makes the model more robust and less prone to overfitting.
- SVMs are robust to noise, as they are less affected by the presence of outliers in the data.
- SVMs can be easily extended to multi-class classification problems using one-vs-one or one-vs-all schemes.

Overall, SVMs are a robust and efficient machine learning technique, widely applied across various domains.

2.5.2 How Does SVM Algorithm Work?

A basic linear SVM classifier divides two classes using a straight line. All data points on one side of the line are assigned to one category, while those on the opposite side are assigned to another. This highlights that an infinite number of lines could be chosen [26]. To better understand how the SVM algorithm works, imagine a dataset with two labels (green and blue) and two features (x1 and x2). The goal of the classifier is to categorize each pair of coordinates (x1, x2) as either green or blue. The following Fig. 5 helps to illustrate this:

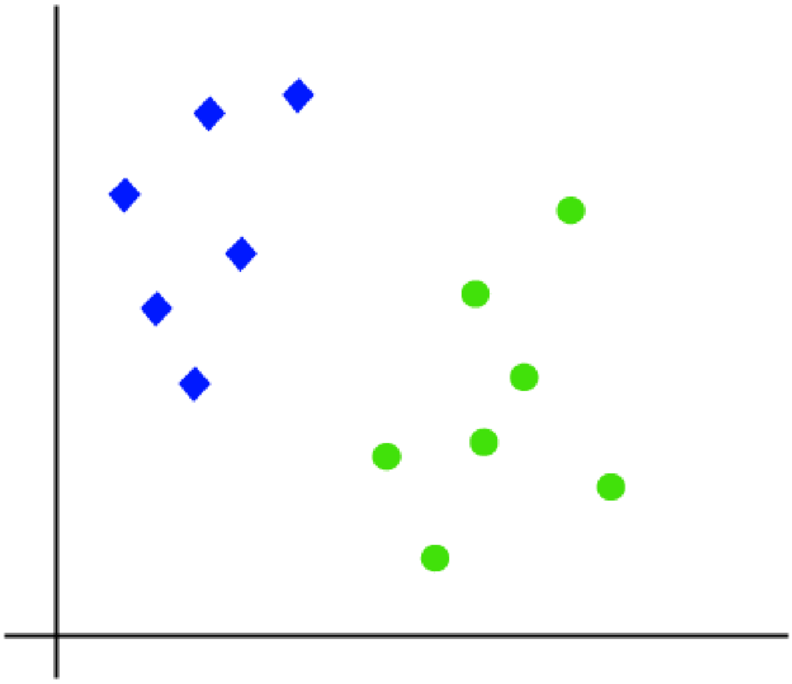


Fig. 5: First step in SVM works

Since this is a 2-D space, it is straightforward to separate the two classes with a straight line. However, there could be several lines that can effectively divide these classes. See Fig. 6 below for an example:

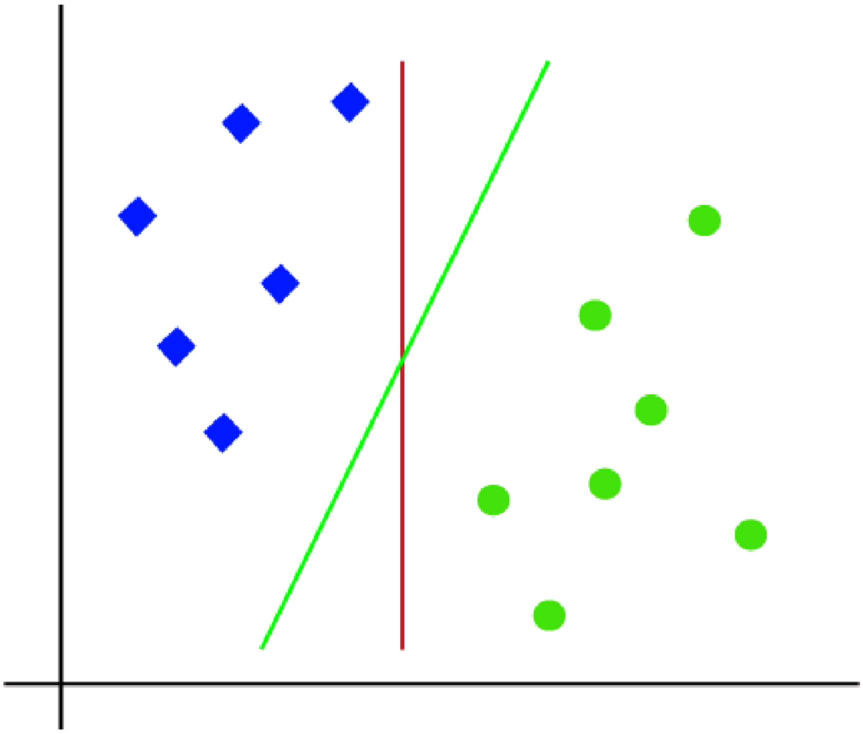


Fig. 6: Second step in SVM works

.
In the SVM algorithm, the optimal decision boundary, called the hyperplane, is determined by locating the support vectors, which are the closest points from each class. The margin is the distance between the support vectors and the hyperplane, and the goal of the SVM is to maximize this margin. As a result, the hyperplane with the largest margin is considered optimal. This is demonstrated in Fig. 7 below:

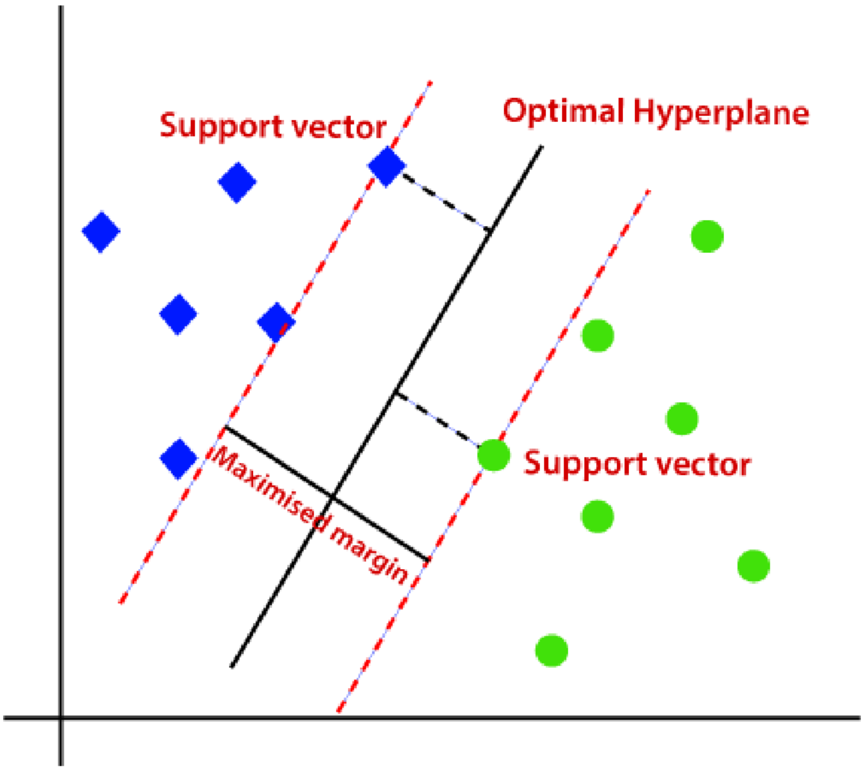


Fig. 7: Result in SVM works

Kernel tricks, also known as the Generalized Dot Product, involve computing the dot product between two vectors to assess how much they influence each other. According to Cover's theorem, the chances of linearly non-separable data becoming separable increase as the dimensionality rises. Kernel functions are used in SVM-constrained optimization to compute these dot products [27].

The SVM kernel is a function that maps low-dimensional input data into a higher-dimensional space, transforming non-separable problems into separable ones. This is particularly useful for solving nonlinear separation problems. In simple terms, the kernel performs complex data transformations and then identifies the separation process based on the labels or outputs specified [24]. Fig. 8 illustrates the SVM Kernel Function.

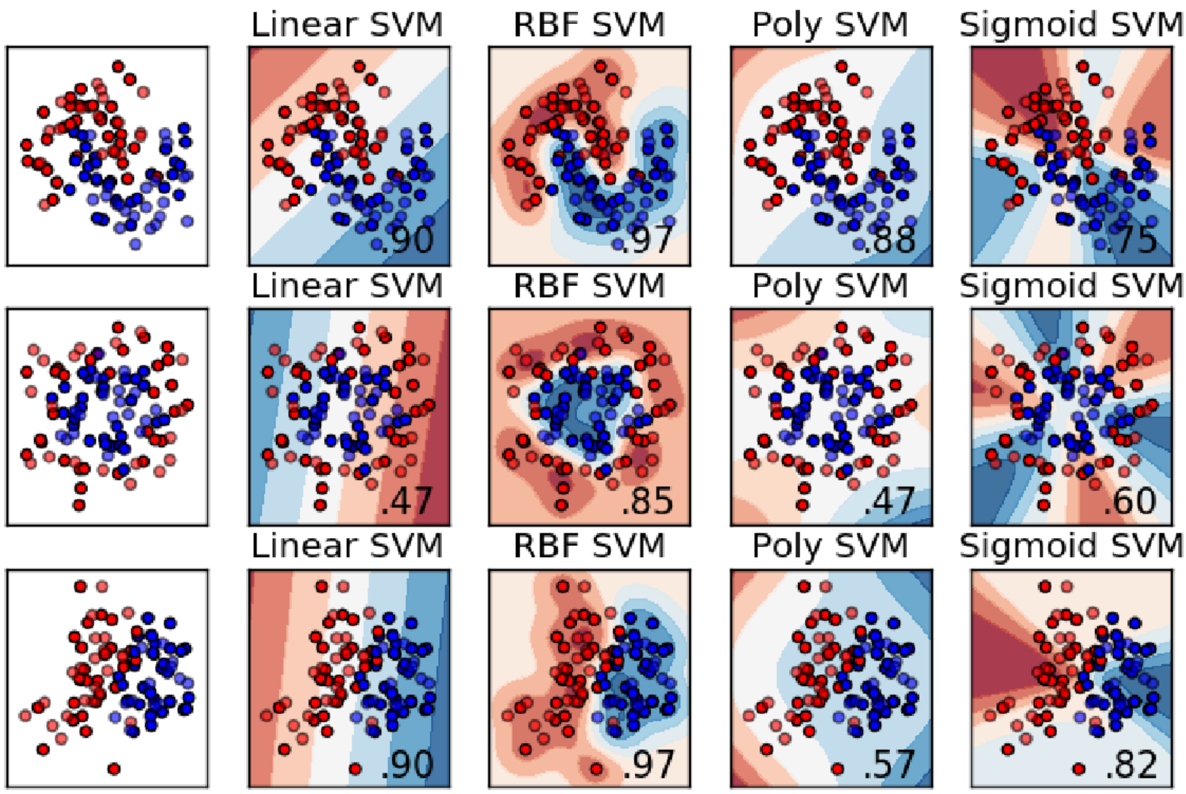


Fig. 8: SVM kernel function

In summary, SVM is a supervised learning method that uses a hyperplane to separate data into distinct classes. The optimal hyperplane is the one that maximizes the margin between itself and all data points, ensuring the best possible classification. While SVM is a highly flexible and memory-efficient technique, it can suffer from overfitting, may become computationally expensive with large datasets, and lacks a probabilistic interpretation of the results [28].

2.5.3 SVM Algorithms Pseudocode

SVMs are a type of supervised machine learning algorithm used for both classification and regression tasks. Below is a basic pseudocode for training an SVM on a classification problem:

1 Load the training data and labels.
2 Initialize the SVM with parameters such as the kernel type and any regularization constants.
3 Pre-process the data, if necessary (e.g., scaling or centering).
4 Select a subset of the training data to use as support vectors. This can be done randomly or by selecting points that are closest to the decision boundary.
5 Train the SVM by optimizing the weights of the support vectors.
6 Test the SVM on the test data and evaluate its performance.

Fig. 9 presents an example of Python code for training and applying an SVM to a classification task.

```
# import necessary libraries
from sklearn.svm import SVC
from sklearn.model_selection import train_test_split

# split data into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

# create an SVM model
model = SVC()

# train the model
model.fit(X_train, y_train)

# make predictions on the test data
y_pred = model.predict(X_test)

# evaluate the model's performance
accuracy = model.score(X_test, y_test)
print("Accuracy:", accuracy)
```

Fig. 9: Example code for train SVM

2.5.3 Flowchart of SVM Algorithms

The SVM process is exemplified in the flowchart shown in Fig. 10 below. This flowchart provides a more specific guide for applying an SVM to classification tasks, highlighting the steps for training, testing, and making predictions:

1. Collect and pre-process the data. This includes cleaning the data, handling missing values, and possibly scaling or normalizing the features.
2. Split the data into training and testing sets.
3. Choose an SVM kernel and hyperparameters. The kernel is a function that determines the decision boundary, and hyperparameters are values that control the behavior of the kernel.
4. Train the SVM model on the training data using the chosen kernel and hyperparameters.
5. Evaluate the model on the testing data. Calculate relevant evaluation metrics, such as accuracy, precision, and recall.
6. Fine-tune the model by adjusting the kernel and hyperparameters, if necessary.
7. Use the trained model to make predictions on new, unseen data.
8. Optionally, retrain the model on the combined training and test data, and make additional predictions as needed.
9. Analyze and interpret the results of the predictions. This may include visualizing the decision boundary, examining the evaluation metrics, and understanding the implications of the predictions for the specific problem.

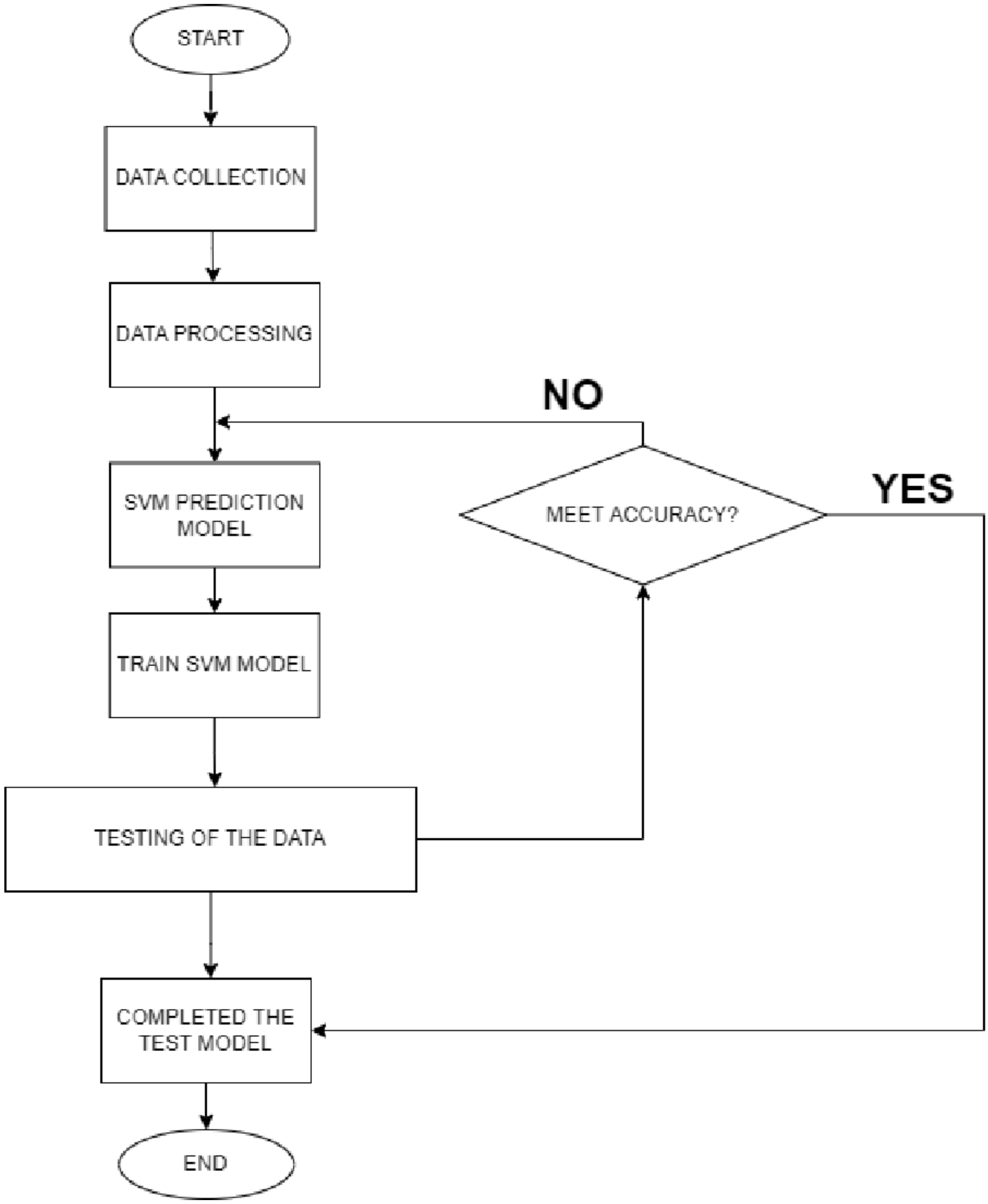


Fig. 10: Example of a flowchart for the SVM process

3.0 RESEARCH METHODOLOGY

3.1 System Architecture and Flowchart

Fig. 11 below illustrates the system architecture. It involves a database for storing user data, a dataset containing temperature, humidity, and rainfall information, and an algorithm based on support vector machines. The entire system is developed as a mobile application.

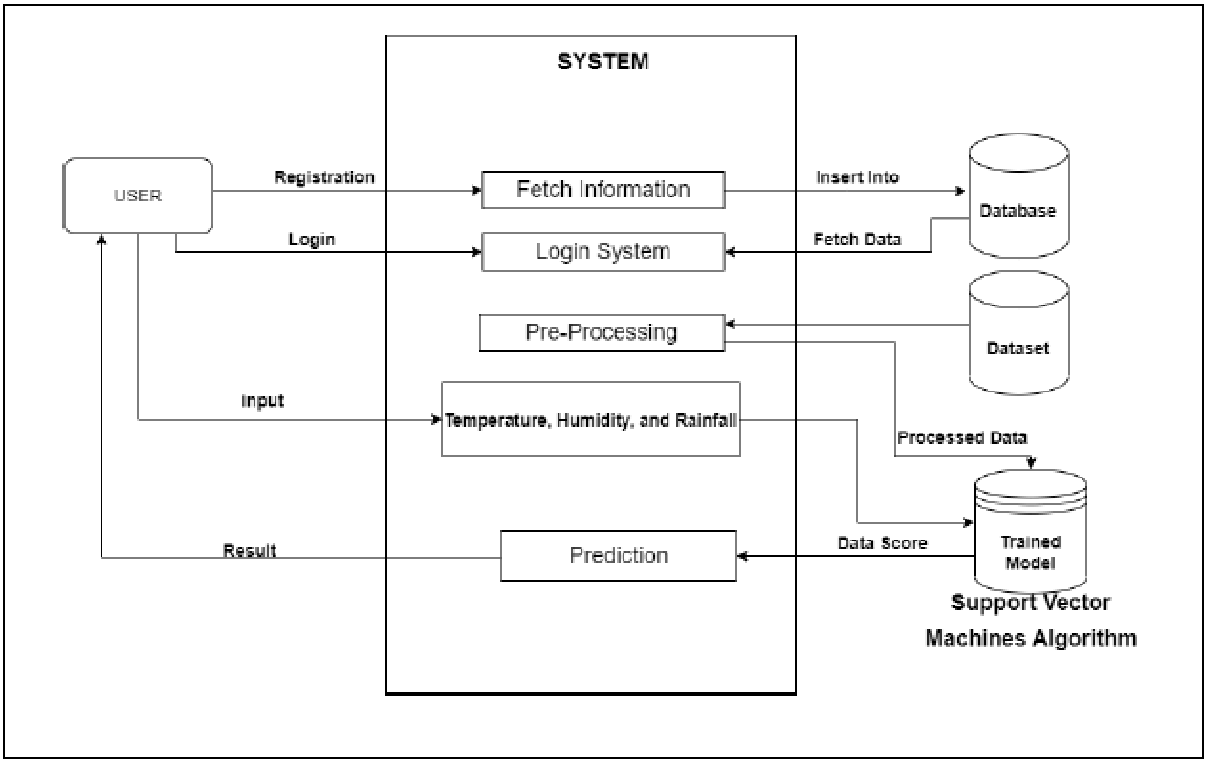


Fig. 11: The system architecture

Fig. 12 shows a flowchart of SVM Algorithm to be implemented in the system.

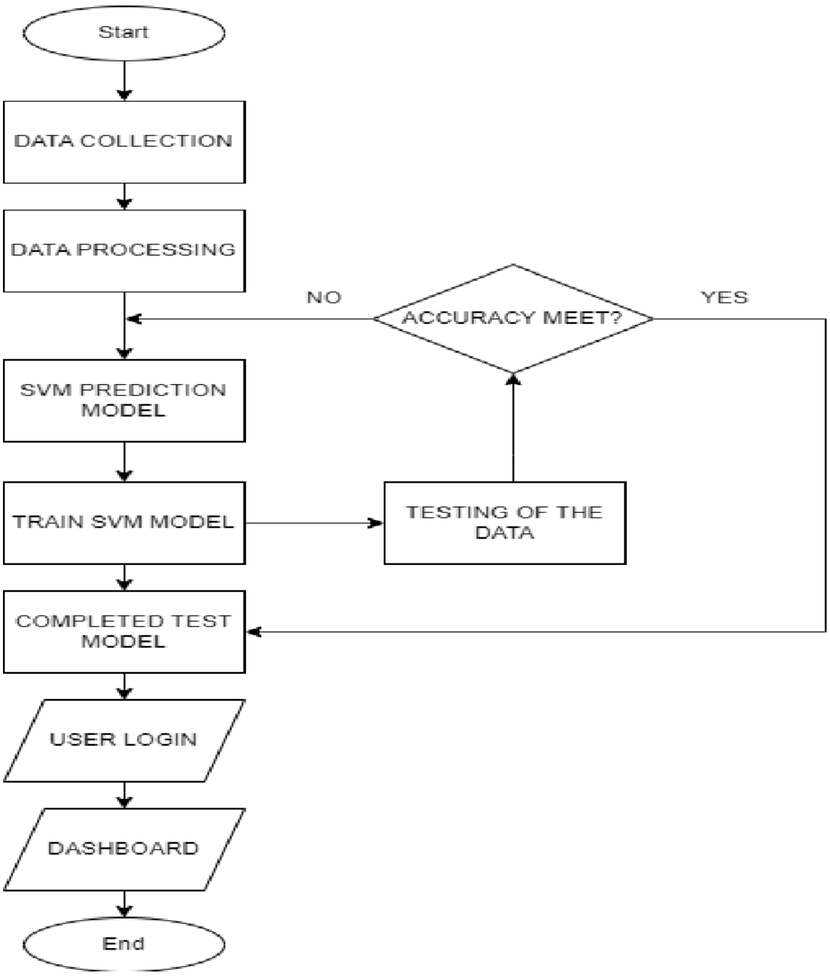


Fig. 12: Flowchart of the system using SVM algorithm

3.2 User Interface Design

User Interface (UI) design involves creating a visually appealing and interactive layout for software or digital products, ensuring it is easy and intuitive for users to navigate and engage with. UI design plays a crucial role in software development, as it shapes how users interact with the product and influences their overall experience. Figures 13 (a), (b), and (c) present the user interface for this system.

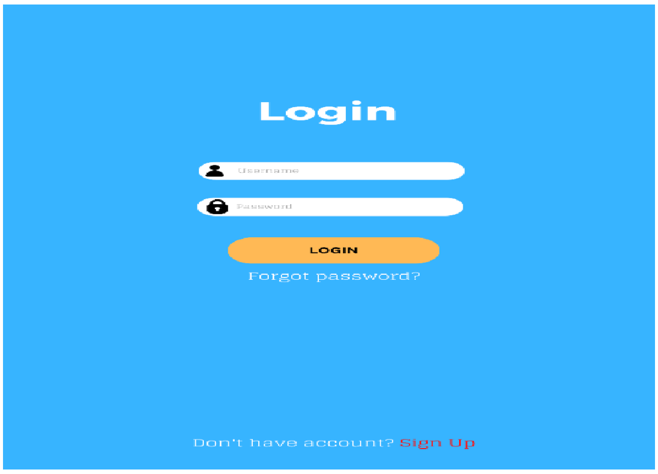


Fig. 13 (a): Login UI

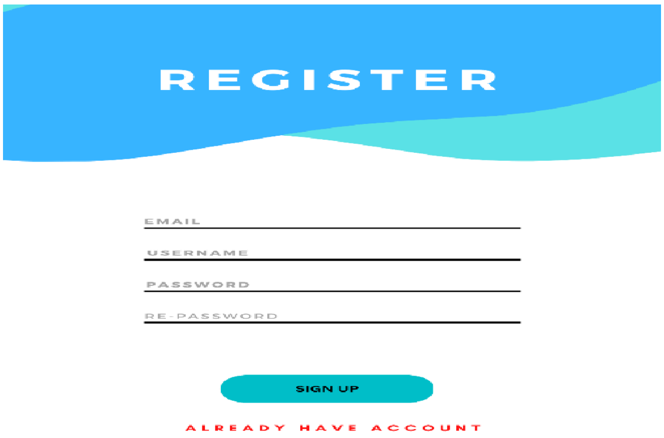


Fig. 13 (b): Register UI

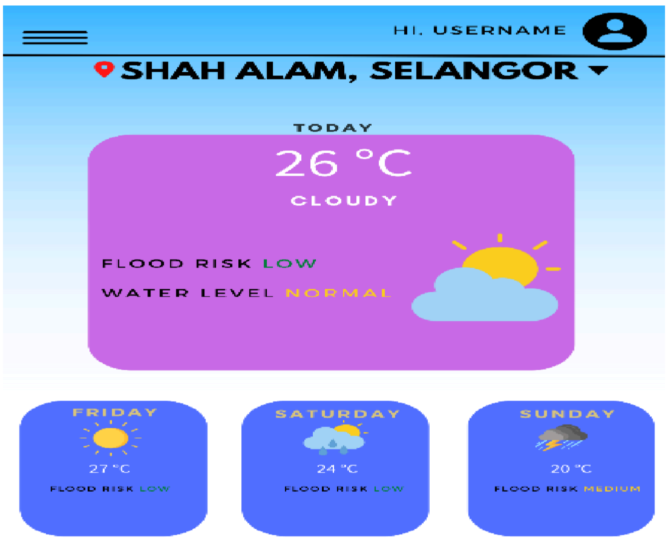


Fig. 13 (c) : Dashboard UI

3.3 Pseudocode of SVM Algorithm

Fig. 14 provides examples of Java code implementing the SVM algorithm for flood prediction, with data sourced from a CSV file.

```
// Import necessary libraries
import java.io.File;
import java.io.IOException;
import weka.core.Instances;
import weka.core.converters.CSVLoader;
import weka.classifiers.Evaluation;
import weka.classifiers.functions.SMO;

public class FloodMonitoring {
    public static void main(String[] args) throws Exception {

        // Load CSV file containing train and test data
        CSVLoader loader = new CSVLoader();
        loader.setSource(new File("flood_data.csv"));
        Instances data = loader.getDataSet();

        // Preprocess the data
        // ...

        // Split the data into train and test sets
        int trainSize = (int) Math.round(data.numInstances() * 0.8);
        int testSize = data.numInstances() - trainSize;
        Instances trainData = new Instances(data, 0, trainSize);
        Instances testData = new Instances(data, trainSize,
testSize);

        // Train the SVM model using the train data
        SMO svm = new SMO();
        svm.buildClassifier(trainData);

        // Test the SVM model using the test data
        Evaluation eval = new Evaluation(testData);
        eval.evaluateModel(svm, testData);
        double accuracy = eval.pctCorrect();

        // Monitor data from sources such as weather stations,
satellite imagery and social media
        // ...

        // Input the monitored data into the trained SVM model
        double[] predictions =
svm.distributionForInstance(monitoredData);

        // Evaluate the results of the predictions
        // ...

        // If the risk of flood is high, send alerts to relevant
authorities and people in the affected area
        if (floodRiskHigh) {
            sendAlerts();
        }
```

```
        // Repeat steps 6 to 10 until the flood risk subsides
        // ...

        // Update the SVM model with new data to improve its accuracy
        svm.updateClassifier(updatedData);
    }

    private static void sendAlerts() {
        // Code to send alerts
    }
}
```

Fig. 14: Coding of SVM algorithm

3.4 Prediction Accuracy

A confusion matrix is a table used to evaluate classification models like the SVM method [29]. It details the model's true positives, true negatives, false positives, and false negatives. Table 1 shows flood prediction model confusion matrix:

Table 1: Flood prediction model confusion matrix

	Predicted Flood	**Predicted No Flood**
Actual Flood	True Positive (TP)	False Negative (FN)
Actual Flood	False Positive (FP)	True Negative (TN)

- True Positive (TP) represents the number of instances where the model correctly predicted a flood and there was a flood.
- True Negative (TN) represents the number of instances where the model correctly predicted no flood and there was no flood.
- False Positive (FP) represents the number of instances where the model predicted a flood but there was no flood.
- False Negative (FN) represents the number of instances where the model predicted no flood but there was a flood.

From the confusion matrix, several evaluation metrics can be calculated such as accuracy, precision, recall, and F1-score.

- Accuracy = (TP + TN) / (TP + TN + FP + FN)
- Precision = TP / (TP + FP)
- Recall = TP / (TP + FN)
- F1-Score = 2 * (Precision * Recall) / (Precision + Recall)

These evaluation metrics provide additional information about the performance of the model and depending on the problem and the goals of the model, different evaluation metrics may be more appropriate. For example, in some situations, it may be more important to minimize false negatives than false positives.

3.5 Error Rate

Error rate, or misclassification rate, is a widely used metric for evaluating the effectiveness of classification models, including the Support Vector Machine (SVM). It reflects the ratio of incorrect predictions to the total number of predictions made. This can be calculated by dividing the number of wrong predictions by the total number of predictions. Here is an example of how the error rate is derived from a confusion matrix as shown in Equation (1):

$$\textit{Error Rate} = (FP + FN) / (TP + TN + FP + FN) \quad (1)$$

Where:
- FP: False Positive
- FN: False Negative
- TP: True Positive
- TN: True Negative

3.6 K-Fold Cross Validation

K-fold cross-validation is the process of dividing a data set into k equal subsets, or "folds," and then training and evaluating the model k times, each time using a different subset as the test set and the remaining k subsets as the training set. The model's performance is then averaged over the k iterations to get a better idea of how well it works.

The formula for k-fold cross-validation varies depending on the chosen evaluation metric. For instance, if accuracy is the evaluation metric, the formula for k-fold cross-validation accuracy is given in Equation 2:

$$(1/k) * \Sigma(i=1 \text{ to } k) (Accuracy\_i) \quad (2)$$

where *Accuracy_i* is the accuracy of the model on the *i-th* iteration.

Similarly, if the evaluation metric is mean squared error (MSE), the formula for k-fold cross-validation MSE is expressed in Equation 3:

$$(1/k) * \Sigma(i=1 \text{ to } k)\ (MSE\_i) \quad (3)$$

where MSE_i is the mean square error of the model on the *i-th* iteration.

4.0 RESULTS AND FINDINGS

4.1 Conceptual Framework

The conceptual framework for flood prediction using SVM algorithms in a mobile application is shown in Fig. 15. It begins with the user logging into the mobile application, which then directs the user to the dashboard. Afterward, the user inputs data on temperature, humidity, and rainfall to predict floods. An Application Programming Interface (API) containing a pre-trained SVM model interacts with a button within the application, displaying the output to the user.

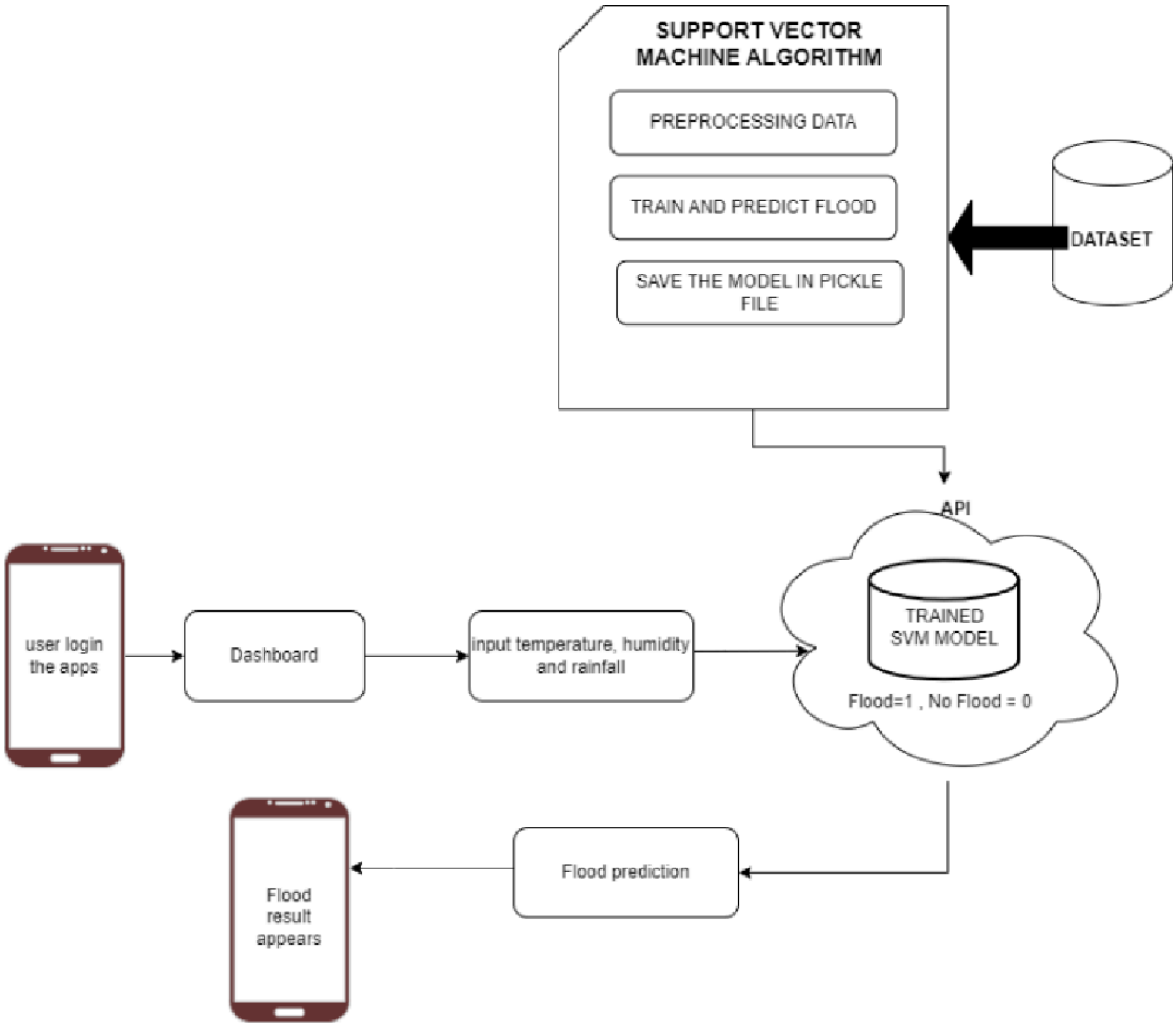
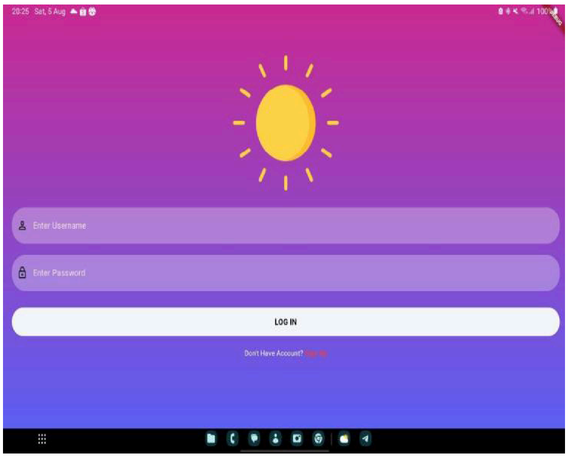
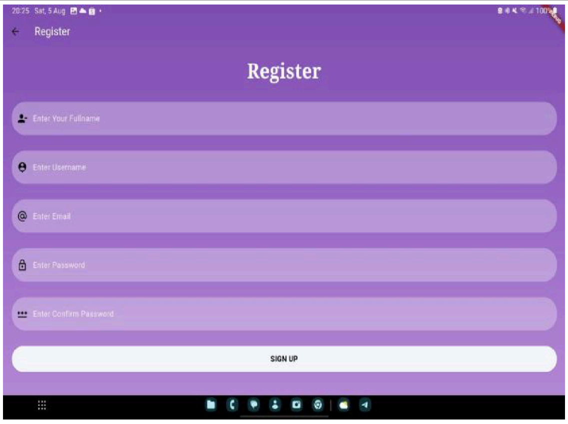
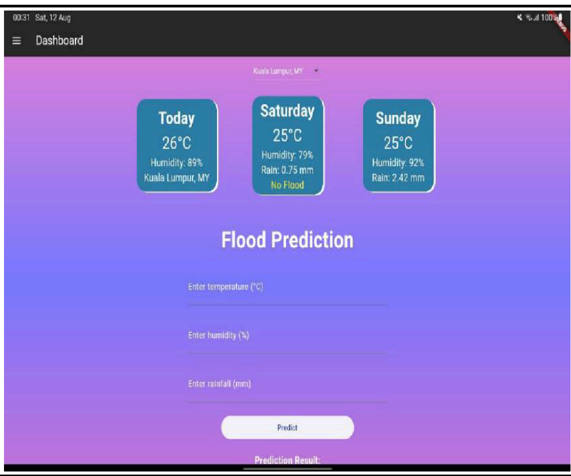


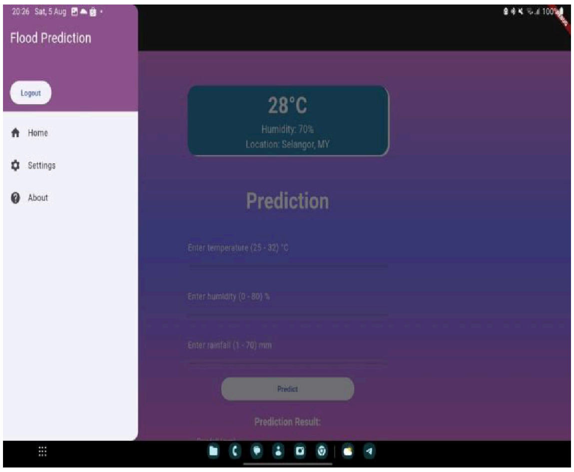
Fig. 15: Conceptual framework

4.2 Prototype of Flood Prediction System using SVM

The dataset used was obtained from the Kaggle database website, comprising a total of 2,421 data points that include temperature, humidity, and rainfall variables for flood prediction. Indicator rainfall data is sourced from *publicinfobanjir.water.gov.my* [30]. The application was successfully developed using the Flutter programming language, enabling a fully functional mobile application. Additionally, Python was utilized to create an API for interacting with the mobile application to facilitate flood predictions. Table 2 displays all the snapshot of the prototype and its description.

Table 2: Snapshot and description of prototype

Snapshot of Prototype	Description
	**Login Page** 1. Login page for users to log into the system. 2. Users input their email and password.
	**Register Page** 1. Register page for users to create an account. 2. Users input full name, username, email, password and confirm password.
	**Dashboard Page** 1. The dashboard page shows the temperature and humidity in Selangor. 2. Weather forecasting is sourced from *OpenWeather.com*, displaying current weather, the next day's forecast and the forecast for the following two days.

 	**Drawer** 1. Provide navigation to the home, settings, and about pages. 2. Add a logout button to allow users to log out of their accounts from the application.

4.3 Evaluation Result

For the evaluation, this study employed three tuning methods: StandardScaler, SMOTE, and GridSearchCV. The description of each method is provided in Table 3.

Table 3: A description of each tuning method

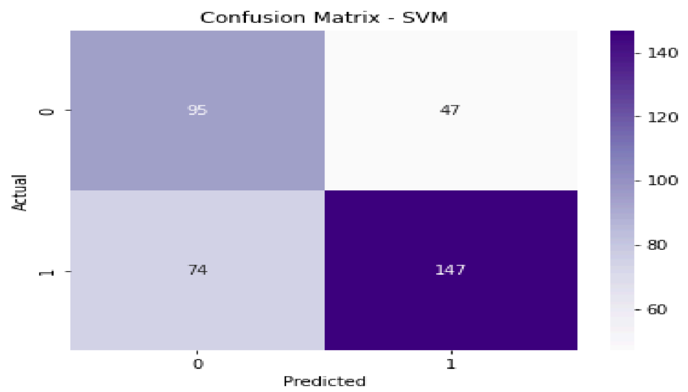
Method	Purpose	Description
StandardScaler	Feature Scaling	StandardScaler is used to scale the features of the dataset to have zero mean and unit variance. It transforms the data so that each feature has a mean of 0 and a standard deviation of 1, ensuring features are on the same scale.
SMOTE	Address Class Imbalance	Synthetic Minority Over-sampling Technique (SMOTE) is used to address the issue of class imbalance in the training data. It creates synthetic samples for the minority class by interpolating between existing samples, effectively balancing the class distribution, and preventing bias towards the majority class during training.
GridSearchCV	Hyperparameter Tuning and Model Selection	GridSearchCV is a hyperparameter tuning technique used to find the best combination of hyperparameters for a given model. It performs an exhaustive search over a specified set of hyperparameter values and evaluates the model's performance using cross-validation. The combination with the best performance is selected as the final model.

4.4 Test Size

For the SVM model, “train_test_spilt” function is used. This function is a utility provided by the scikit-learn library in Python, and is commonly used in machine learning tasks for model evaluation and testing. Its purpose is to split a dataset into two subsets; the training set and the test set. Table 4 shows the test results for test size.

Table 4: Result of testing

<table>
<tr><td>Test and Train Size</td><td>Ratio – Testing : Training (1.5 : 8.5)<pre>from sklearn.model_selection import train_test_split
from imblearn.over_sampling import SMOTE
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.15, random_state=42)</pre></td></tr>
</table>

Testing and Training Score	Accuracy (SVM): 66.66666666666666 Accuracy on train (SVM): 63.68464052287581
Confusion Matrix	
Evaluation Metrics	Accuracy = ((95+147)/(95+47+74+147))*100 = 66.66 %
Classification Report	Classification Report: precision recall f1-score support 0 0.56 0.67 0.61 142 1 0.76 0.67 0.71 221 accuracy 0.67 363 macro avg 0.66 0.67 0.66 363 weighted avg 0.68 0.67 0.67 363

Based on Table 4, this prototype achieves the highest accuracy for testing, which is 66.66%. The ratio of testing to training for achieving this highest accuracy is 15 for testing and 85 for training.

Table 5: Summary of all testing

Evaluation Measures	Training and Testing Split Ratio (Test : Train)		
	Experiment 1: (10:90)	Experiment 2: (15:85)	Experiment 3: (20:80)
Confusion Matrix	[52 33] [49 108]	[95 47] [74 147]	[121 68] [105 190]
Accuracy	66.11%	66.66%	64.25%

Three different test sizes were used to determine the accuracy and percentage error of the model while other parameters were fixed throughout the process. Table 5 summarizes all the testing for each test size. The evaluation involved calculating various metrics such as accuracy, confusion matrix, precision, recall, and F1-score. Based on the result obtained, it can be concluded that the entire prototype of the flood prediction system using SVM on a mobile application functions effectively, successfully completing all tasks in the functional test.

SVMs are a type of machine learning algorithm used for classification and regression tasks. They work by finding the hyperplane that best separates the data into two classes. In the context of flood prediction, these classes would typically be "flood" and "no flood. SVMs have shown effectiveness in flood prediction by learning patterns in data associated with floods. For instance, an SVM could be trained on historical flood data in Malaysia to recognize relationships between variables like temperature, humidity, and rainfall, enabling it to predict flood likelihood. Although the use of SVMs for flood prediction is still in its early stages, but it has the potential to become a powerful tool for enhancing our understanding of floods and improving response strategies.

5.0 CONCLUSION

This study aims to create a mobile application for flood prediction using the SVM algorithm. To utilize the SVM for flood prediction in the mobile app, data collection and preprocessing were essential steps. The dataset used in this study, which includes 2,421 data points with temperature, humidity, and rainfall variables, was sourced from the Kaggle database. Rainfall data was provided by publicinfobanjir.water.gov.my.

The application was successfully developed using the Flutter programming language, resulting in a fully operational mobile application. Python was also used to create an Application Programming Interface (API) that facilitates communication between the mobile app and the flood prediction system. Extensive functionality testing was conducted, producing positive outcomes, and the majority of users expressed satisfaction with the application.

The evaluation of the accuracy of the SVM algorithm was accomplished by calculating the model's accuracy and F1 score. The prediction accuracy achieved by the SVM algorithm was around 66.66%, which is relatively low for machine learning models, primarily due to the limitations of the dataset. The decision to use the SVM algorithm was based on its ability to effectively handle smaller datasets, making it a suitable choice for this study.

The study aims to make a meaningful contribution to individuals, whether they have experienced flooding or not. By developing a mobile application for flood prediction, users will have access to a portable and convenient tool that can be used anywhere and checked at any time. This mobile application will not only benefit flood-affected individuals but also contribute to flood-related research and assist Malaysia's civil defence force in their duties. The capabilities of the mobile application will help the civil defence force to carry out evacuation procedures more efficiently and simplify their tasks during flood events. By leveraging the support of this predictive system, the civil defence force can make informed decisions and respond promptly to potential flood situations, ensuring the safety and well-being of flood victims.

Additionally, the strategic choice to develop an Android mobile considers the widespread use of Android phones among the population. By adopting this platform, the flood prediction system can reach a larger user base, enabling more individuals to access flood forecasts and prepare for potential flood incidents. This contribution is expected to empower communities with valuable flood prediction insights and improve overall flood preparedness across Malaysia.

There are several suggestions for future system enhancements. Implementing IoT technology can enable real-time data collection, thereby improving prediction accuracy and allowing for timely flood alerts to be sent to users. Hybridizing the application for both Android and iPhone operating systems will expand its reach and utility. The accuracy of the SVM model can be continually enhanced by storing real-time flood parameter data in a database. This mobile application has the potential to serve as a valuable tool for flood prediction and preparedness, benefiting flood-prone communities and enhancing flood management efforts in Malaysia. With the suggested enhancements, the application can effectively meet the needs of its users and improve its contributions.

REFERENCES

[1] L. M. Sidek, H. Basri, M. H. Mohammed, M. Marufuzzaman, N. A. Ishak, A. M. Ishak, B. Z. C. Omar, S. Osman, S. Ramly, and M. H. Hassan, "Towards impact-based flood forecasting and warning in Malaysia: A case study at Kelantan River", IOP Conference Series: Earth and Environmental Science, Vol. 704(1), 012001, 2021.

[2] D. S. DR. M. U. Mahidin, Department of statistics Malaysia official portal, *www.dosm.gov.my*, 2022.

[3] N. Theera-Umpon, S. Auephanwiriyakul, S. Suteepohnwiroj, and J. Pahasa, "River basin flood prediction using support vector machines", 2020.

[4] E. Yusuf, H. Yuk Feng, N. Jing Lin, H. B. Ahmed, A. Ali Najah, and, E.-S. Ahmed, "Predicting streamflow in peninsular Malaysia using support vector machine and deep learning algorithms", Scientific Reports, 2022.

[5] T. Khan, M. Alam, P. Karachi, P. Su'ud, and K. Kadir, "Early flood risk assessment using machine learning: A comparative study of SVM, Q-SVM, K- NN and LDA", Zeeshan Shahid Institute of Business Management (IOBM), In Computer Science and Statistics, MACS, 2019.

[6] J. M. Antony Sylvia, M. Pushpa Rani, and B. Aremu, "Analysis of IoT big weather data for early flood forecasting system", Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2021.

[7] S. Agarwal, P. J. Roy, P. Choudhury, and N. Debbarma, "Flood forecasting and flood flow modeling in a river system using ANN", Water Practice and Technology, Vol. 16(4), pp. 1194–1205, 2021.

[8] S. Mekruksavanich, K. Sooksomsatarn, and A. Jitpattanakul, "Flooding forecasting system based on water monitoring with IoT technology", IEEE 12th International Conference on Software Engineering and Service Science (ICSESS), 2021.

[9] M. M. A. Syeed, M. Farzana, I. Namir, I. Ishrar, M. H. Nushra, and T. Rahman, "Flood prediction using machine learning models", International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2022.

[10] F. Jerome Glago, "Flood disaster hazards; causes, impacts and management: A state-of-the-art review", Natural Hazards - Impacts, Adjustments and Resilience, Vol. 1(1), 2021.

[11] Ausmed Education Pty Ltd, "Predicting the future: What is health forecasting? | ausmed", Www.ausmed.com, *https://www.ausmed.com/publish/handover/articles/health-forecasting*, May 2022.

[12] StudySmarter, "Coastal flooding: Definition, causes & solution | studysmarter, StudySmarter UK, *https://www.studysmarter.co.uk/explanations/geography/coasts-geography/coastal-flooding/*, 2022.

[13] N. Bradford, "Increases in coastal flooding", NEEF, *https://www.neefusa.org/nature/land/increases-coastal-flooding*, 2022.

[14] World Health Organization, "Floods", Www.who.int, *https://www.who.int/health-topics/floods#tab=tab_1*, 2021.

[15] Internet Geography, "How does flooding affect humans and the environment?" Internet Geography; cre8tiveit.solutions, *https://www.internetgeography.net/topics/how-does-flooding-affect-humans-and-the-environment/*, 2022.

[16] R. Ali, "The why behind financial forecasting", Oracle NetSuite, *https://www.netsuite.com/portal/resource/articles/financial-management/importance-financial-forecasting.shtml*, October 2020.

[17] Justin, "Impact of Floods on the Environment and Humans – Globalenergy-news.com", Impact of Floods on the Environment and Humans; Justin. *https://globalenergy-news.com/impact-of-floods-on-the-environment-and-humans/*, August 2020.

[18] B. Eren, "Floods and Their Impact on the Environment", *https://www.researchgate.net/publication/356257409_Floods_and_Their_Impact_on_the_Environment*, November 2021.

[19] O'Reilly Media, Inc., 1. introduction - effective monitoring and alerting [book], Www.oreilly.com; Oreilly, *https://www.oreilly.com/library/view/effective-monitoring-and/9781449333515/ch01.html*, 2023.

[20] Vocabulary.com, "Prediction - definition, meaning & synonyms", Vocabulary.com. *https://www.vocabulary.com/dictionary/prediction#:~:text=A%20prediction%20is%20what%20someone*, 2023.

[21] U.S. Department of the Interior, "How are floods predicted?", Usgs.gov, *https://www.usgs.gov/faqs/how-are-floods-predicted*, 2018.

[22] M. Abdul Majid, M. Hafidz Omar, M. S. M. Noorani, and F. Abdul Razak, "River-flood forecasting methods: The context of the kelantan river in Malaysia", IOP Conference Series: Earth and Environmental Science, Vol. 880(1), 012021, 2021.

[23] B. Gaye, D. Zhang, and A. Wulamu, "Improvement of Support Vector Machine Algorithm in Big Data Background", Mathematical Problems in Engineering (Special Issue), e5594899, 2021.

[24] A. Sasidharan, "Support vector machine algorithm", GeeksforGeeks, *https://www.geeksforgeeks.org/support-vector-machine-algorithm/*, January 2021.

[25] JavaTpoint, “Support vector machine (SVM) algorithm – javatpoint”, Www.javatpoint.com, *https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm*, 2021.

[26] M. McGregor, “SVM machine learning tutorial - what is the support vector machine algorithm, explained with code examples”, FreeCodeCamp.org, *https://www.freecodecamp.org/news/svm-machine-learning-tutorial-what-is-the-support-vector-machine-algorithm-explained-with-code-examples/*, July 2020.

[27] R. Dwivedi, “How does support vector machine algorithm works in machine learning? | analytics steps”, Www.analyticssteps.com; Analyticssteps, *https://www.analyticssteps.com/blogs/how-does-support-vector-machine-algorithm-works-machine-learning*, January 2021.

[28] Vatsal, “Support vector machine (SVM) explained. Support Vector Machine (SVM) Explained”, *https://towardsdatascience.com/support-vector-machine-svm-explained-58e59708cae3*, May 2022.

[29] J. Yan, J. Jin, F. Chen, G. Yu, H. Yin, and W. Wang, “Urban flash flood forecast using support vector machine and numerical simulation”, Journal of Hydroinformatics, Vol. 20(1), pp. 221–231, 2018.

[30] The Official Web of Public Infobanjir for River Water Level Data, *https://publicinfobanjir.water.gov.my*, 2019.